

Current Molecular Epidemiology of Lassa Virus in Nigeria^{▽§}

Deborah U. Ehichioya,^{1,2,3} Meike Hass,² Beate Becker-Ziaja,² Jacqueline Ehimuan,⁴ Danny A. Asogun,⁴ Elisabeth Fichet-Calvet,² Katja Kleinstaub,² Michaela Lelke,² Jan ter Meulen,^{5†} George O. Akpede,⁴ Sunday A. Omilabu,¹ Stephan Günther,^{2*} and Stephan Ölschläger²

Department of Medical Microbiology and Parasitology, College of Medicine, University of Lagos, Lagos, Nigeria¹; Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany²; Redeemer's University, Ogun State, Nigeria³; Irrua Specialist Teaching Hospital, Irrua, Edo State, Nigeria⁴; and Institute of Virology, Philipps University Marburg, Marburg, Germany⁵

Received 16 September 2010/Returned for modification 12 November 2010/Accepted 20 December 2010

Recent Lassa virus strains from Nigeria were completely or partially sequenced. Phylogenetic analysis revealed the predominance of lineage II and III strains, the existence of a previously undescribed (sub)lineage in Nigeria, and the directional spread of virus in the southern part of the country. The Bayesian analysis also provided estimates for divergence times within the Lassa virus clade.

Lassa virus belongs to the Old World complex of the family *Arenaviridae*. It causes hemorrhagic fever in humans. The disease is endemic in Sierra Leone, Guinea, Liberia, and Nigeria (3, 6). Cases of Lassa fever imported to Europe indicate that the virus also circulates in Ivory Coast and Mali (2, 9). The natural host of Lassa virus is the African rodent *Mastomys natalensis*, which lives close to human settlements (10). Lassa virus may also be transmitted from human to human, giving rise to nosocomial or community-based outbreaks (7).

Lassa virus has a bisegmented genome: the 3.5-kb S RNA codes for glycoprotein (GP) and nucleoprotein (NP), and the 7-kb L RNA codes for large (L) and Z protein. Sequence information for the S RNA of Lassa virus has accumulated in recent years and revealed the presence of four major lineages in West Africa: three in Nigeria (lineages I, II, and III) and one in the area comprising Ivory Coast, Sierra Leone, Liberia, and Guinea (lineage IV) (3).

Although Nigeria is home to the greatest diversity of Lassa virus, the circulating strains are poorly characterized. Only five full-length S RNA sequences are known: for strain LP of lineage I; strain 803213 of lineage II; and strains GA391, Weller, and CSF of lineage III. For the L RNA, there is only one full-length sequence known, that of strain CSF. To fill this gap, we completely sequenced S and L RNA of six Lassa virus strains recently isolated from different areas of Nigeria (5, 13). The full-length sequences as well as a large number of novel partial sequences were subjected to phylogenetic analysis. In addition, a virus isolate from Guinea was completely sequenced and included in the analysis (10).

The origin of the viruses is given in Table 1. Isolates were propagated in the biosafety level 4 laboratory of the Bernhard Nocht Institute. Vero cells were inoculated at a low multiplicity of infection. After 4 days, the supernatant was cleared from the

cell debris and virus was pelleted from the supernatant by ultracentrifugation. RNA was isolated from the pellets by using the QIAamp viral RNA minikit (Qiagen). Overlapping fragments of up to 1,500 nucleotides in length were amplified using various combinations of pan-Old World arenavirus primers targeting conserved sites of S and L RNA and the Qiagen OneStep RT-PCR kit (Qiagen) (see the supplemental material). PCR fragments were sequenced on both strands, and S and L RNA sequences (excluding the conserved 19 nucleotides at the termini) were assembled using SeqMan software (DNASTAR). The remaining gaps were closed after new sets of strain-specific primers were designed. The mean coverage was 3.2 sequence reads per base (range, 2 to 6; on average, 21 reads per S RNA and 38 reads per L RNA). Sequence ambiguities were observed at only two positions (see Table S1 in the supplemental material). Functionally important sequence motifs, such as the GP1/GP2 cleavage site, and endonuclease and polymerase motifs in the L protein are conserved in the novel sequences. The PTAP late domain motif in the Z protein was changed to PSAP in four of the strains (see Table S2 in the supplemental material). If virus could not be isolated in cell culture, Lassa virus genome fragments were amplified from serum samples of patients by using diagnostic reverse transcription (RT)-PCRs targeting the GP or L gene (12, 15) and sequenced.

Phylogenetic analysis included the novel sequences as well as all Old World arenavirus sequences available from GenBank by March 2010. Amino acid sequences of GP, NP, and L genes were aligned by using MacVector software (MacVector) and refined manually. The corresponding nucleotide sequences were aligned manually, guided by the amino acid alignment. Recombination events within genes were not detected by RDP3 software (11). FindModel (<http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html>) identified the general time-reversible model of sequence evolution with a gamma distribution of among-site nucleotide substitution rate variation (GTR+gamma) as the substitution model that best describes the data in the alignments; it was used in all phylogenetic analyses. The fraction of invariant sites was not considered because estimates for this parameter are very sensitive to the number of taxa. Phylogenetic trees were inferred by the

* Corresponding author. Mailing address: Department of Virology, Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Str. 74, 20359 Hamburg, Germany. Phone: 49 40 42818 930. Fax: 49 40 42818 941. E-mail: guenther@bni.uni-hamburg.de.

† Present address: Merck Research Laboratories, West Point, PA.

§ Supplemental material for this article may be found at <http://jcm.asm.org/>.

[▽] Published ahead of print on 29 December 2010.

TABLE 1. Lassa virus strains examined in this study^a

Strain	GenBank accession no.		Specimen	Host	Yr and origin ^b	Patient outcome (reference)
	GP/NP	L				
Nig08-04	GU481068 ^c	GU481069 ^c	I	H	2008, Nigeria, Abakaliki	Died (5)
Nig08-A18	GU481070 ^c	GU481071 ^c	I	H	2008, Nigeria, Jos	Died (5)
Nig08-A19	GU481072 ^c	GU481073 ^c	I	H	2008, Nigeria, Jos	Survived (5)
Nig08-A37	GU481074 ^c	GU481075 ^c	I	H	2008, Nigeria, Irrua (Irrua)	Died
Nig08-A41	GU481076 ^c	GU481077 ^c	I	H	2008, Nigeria, Irrua (Irrua)	Survived
Nig08-A47	GU481078 ^c	GU481079 ^c	I	H	2008, Nigeria, Irrua (Uromi)	Survived
Nig05-043	GU481056	GU481057	S	H	2005, Nigeria, Abakaliki	Died (5)
Nig05-SE40	GU481058	GU481059	S	H	2005, Nigeria, Abakaliki	Survived (5)
Nig07-05	GU481060/1	GU481062	S	H	2007, Nigeria, Jos	Died (5)
Nig08-02	GU481063/4	GU481065	S	H	2008, Nigeria, Abuja	Died (5)
Nig08-03	GU481066	GU481067	S	H	2008, Nigeria, Abakaliki	Died (5)
Nig05-A08	HM143866		S	H	2005, Nigeria, Irrua (Ekpoma)	Survived
Nig07-A09	HM143867		S	H	2007, Nigeria, Irrua (Ekpoma)	Survived
Nig07-A14	HM143868		S	H	2007, Nigeria, Irrua	Died
Nig08-A34	HM143870		S	H	2008, Nigeria, Irrua (Irrua)	Survived
Nig08-A40	HM143871		S	H	2008, Nigeria, Irrua (Uromi)	Survived
Nig08-A45	HM143872		S	H	2008, Nigeria, Irrua (Ekpoma)	Died
Nig08-A53	HM143873		S	H	2008, Nigeria, Irrua (Irukep)	Died
Nig08-A55	HM143874		S	H	2008, Nigeria, Irrua (Irrua)	Survived
Nig08-A57	HM143875		S	H	2008, Nigeria, Irrua (Lokoja)	Survived
Nig08-A61	HM143876		S	H	2008, Nigeria, Irrua (Irrua)	Survived
Nig08-A64	HM143877		S	H	2008, Nigeria, Irrua (Ekpoma)	Died
Nig08-A72	HM143878		S	H	2008, Nigeria, Irrua (Uromi)	Survived
Nig07-A76	HM143879		S	H	2007, Nigeria, Irrua (Uromi)	Survived
Nig08-A77	HM143880		S	H	2008, Nigeria, Irrua (Ekpoma)	Survived
Nig08-A80	HM143881		S	H	2008, Nigeria, Irrua (Otu)	Died
BA366	GU830839 ^c	GU979513 ^c	I	R	2003, Guinea, Bantou	NA (10)

^a Abbreviations: GP, glycoprotein gene; NP, nucleoprotein gene; L, L gene; I, cell culture isolate; S, serum of patient; H, human; R, rodent (*Mastomys natalensis*); NA, not applicable.

^b Location of hospital where patient was treated or sampling site for rodent. If known, the hometown of patient is given in parentheses.

^c Complete gene sequence.

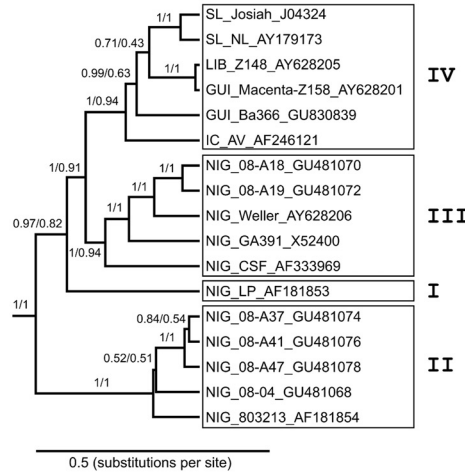
Bayesian Markov Chain Monte Carlo method implemented in BEAST software (4). Complete gene sequences were analyzed under the assumption of a relaxed molecular clock with a fixed substitution rate. The tree topology was verified by a maximum likelihood approach and bootstrap testing implemented in the PhyML program (8). As partial gene sequences were available for a large number of Lassa virus strains isolated over a period of 40 years, the respective alignments contained sufficient temporal structure to estimate substitution rates and divergence times. They were analyzed with BEAST under the assumption

of a strict molecular clock with a substitution rate estimated from the data sets.

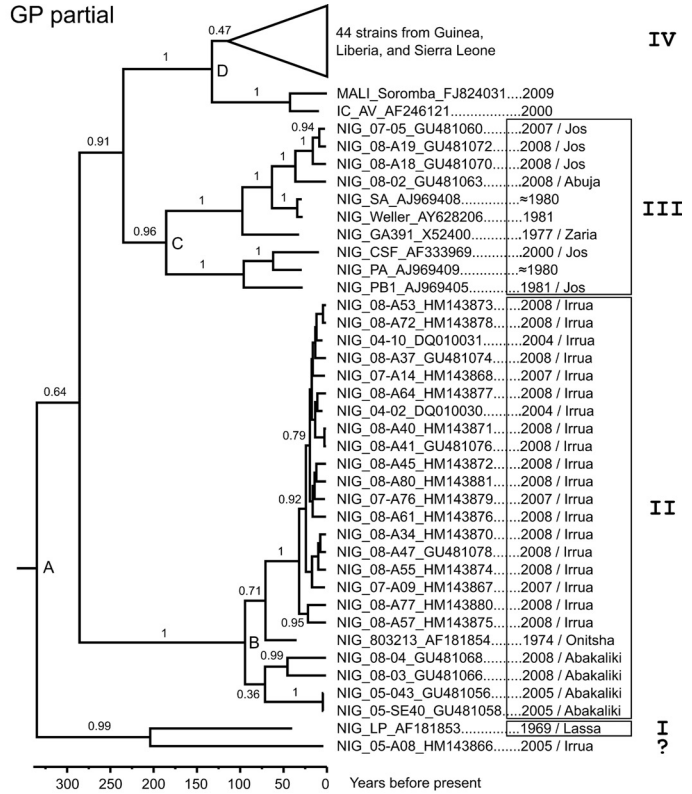
Phylogenetic analysis of complete GP, NP, and L gene sequences confirmed the existence of four Lassa virus lineages (3) (Fig. 1, left; see Fig. S1 in the supplemental material). In particular, the NP and L gene-based trees were well supported by posterior probability (all branches ≥ 0.99) and bootstrap values (≥ 0.94 , with the exception of three branches with values in the range 0.56 to 0.85). In GP-based phylogeny, lineage II is basal, followed by lineages I, III, and IV, while in NP-based

FIG. 1. Phylogenetic analysis of Lassa virus. (Left panel) Phylogenies were inferred using complete nucleotide sequences of GP (1,473 nucleotides), NP (1,707 nucleotides), and L genes (6,654 nucleotides) of all Old World arenaviruses. For clarity of presentation, only the Lassa virus clade is shown; the complete trees are shown in Fig. S1 in the supplemental material. Analysis was performed with BEAST software with the following parameters: general time-reversible model with gamma-distributed sites (GTR+gamma), relaxed lognormal clock with mean substitution rate fixed at 1, 10^7 steps with sampling every 10^5 steps, and results of two independent runs combined (effective sampling size, >200 for all parameters). The data sets were also analyzed using PhyML software with the following parameters: GTR+gamma and consensus of 500 bootstrap trees (not shown). The topology of Bayesian and PhyML trees was congruent for a given gene. BEAST and PhyML support values are indicated on the branches (posterior/bootstrap). (Right panel) Phylogenies were inferred using partial nucleotide sequences of GP (237 nucleotides), NP (631 nucleotides), and L genes (342 nucleotides) of all Lassa virus strains. For clarity of presentation, only Nigerian Lassa virus strains are shown; the complete trees are shown in Fig. S2 in the supplemental material. Analysis was performed with BEAST software with the following parameters: GTR+gamma, strict clock, Bayesian skyline demographic model, 10^7 steps with sampling every 10^5 steps, and results of two independent runs combined (effective sampling size, >200 for all parameters). Trees obtained with an exponential growth model were essentially identical (not shown). Posterior values are indicated on the branches. Nodes representing the most recent common ancestors of the Lassa virus clade and lineages II, III, and IV are marked with letters A to D, respectively. GenBank accession number and year and place of isolation are shown with the strains. The origin of Lassa virus strains is indicated by a prefix: SL, Sierra Leone; LIB, Liberia; GUI, Guinea; IC, Ivory Coast; NIG, Nigeria. Lassa virus lineages are indicated with each tree.

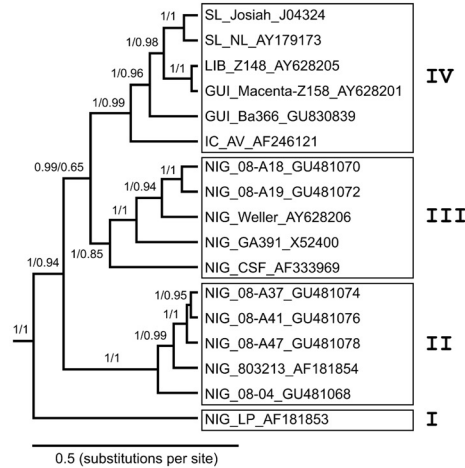
GP complete



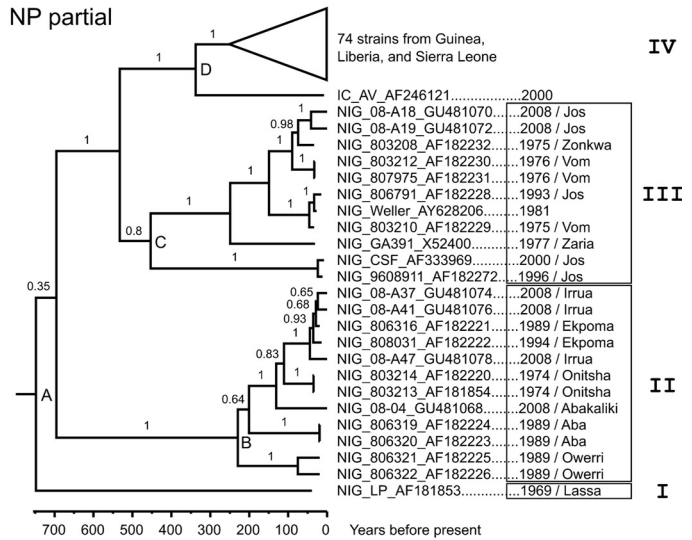
GP partial



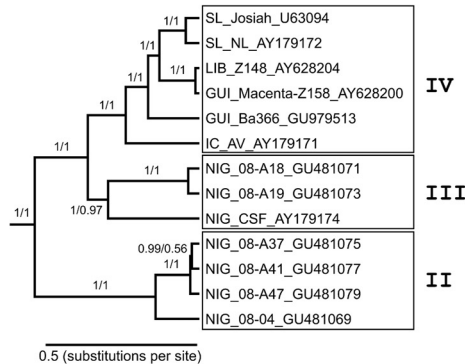
NP complete



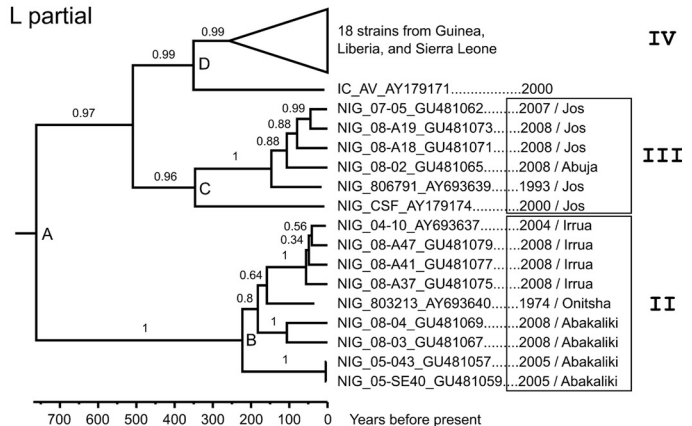
NP partial



L complete



L partial



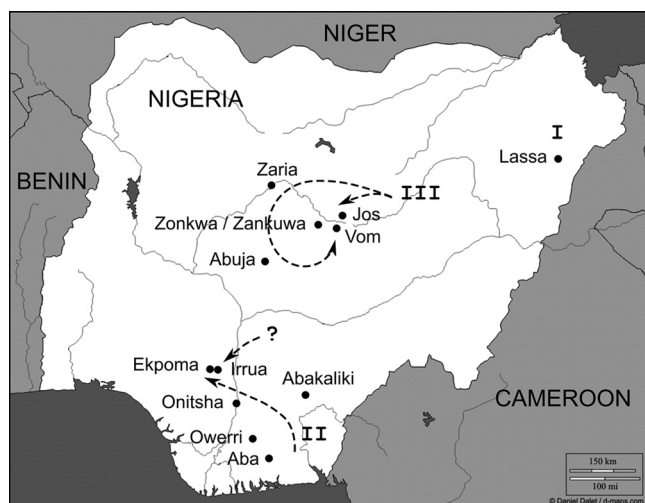


FIG. 2. Map of Nigeria showing sites of Lassa virus circulation. The possible spread of virus is indicated by dashed arrows. The Lassa virus lineages are indicated with the corresponding areas in which the virus is endemic. The potential lineage/sublineage represented by strain Nig05-A08 is indicated with a question mark. (Modified from a map by Daniel Dalet that is freely available at <http://d-maps.com>.)

phylogeny, lineage I is basal, followed by lineages II, III, and IV. This suggests that lineage I may have undergone recombination between NP and GP genes during evolution. However, analysis of the complete S RNA sequences by RDP3 software (11) did not reveal evidence for recombination. Thus, the reason for the ambiguous position of lineage I is unclear. The novel sequence from Guinea (Ba366) clusters in all trees with lineage IV and branches between strain AV from Ivory Coast and the Liberian/Guinean strains Z148 and Z158. The topology within lineage IV is well supported in NP and L gene trees (posterior values, 1.0; bootstrap values, ≥ 0.96). Thus, it appears that, within the western part of West Africa, the virus spread from Ivory Coast via Guinea and Liberia to Sierra Leone.

The novel Nigerian sequences from the southern part of the country (Nig08-A37, Nig08-A41, and Nig08-A47 from Irrua and Nig08-04 from Abakaliki) cluster with lineage II, while those from the central part (Nig08-A18 and Nig08-A19 from

Jos) cluster with lineage III in all three trees. This indicates a geographical clustering of the Nigerian strains, consistent with a previous analysis of partial NP gene sequences (3). To search for geographical and temporal patterns in the phylogeny of Nigerian strains in more detail, a large number of partial GP, NP, and L gene sequences of Lassa virus were subjected to phylogenetic analysis using BEAST, including estimation of substitution rates and divergence times (Fig. 1, right; see Fig. S2 in the supplemental material). Even though the length of the partial sequences was short, the topology of the trees fairly well reflected the topology of the trees inferred with complete gene sequences. All but two of the Nigerian sequences clustered with lineages II and III in all three trees. The two remaining strains were LP, the prototypic strain of lineage I from the town of Lassa, and strain Nig05-A08, originating from a patient who had been treated in Irrua. Unfortunately, the virus load was low, so only the sequence of the fragment of the diagnostic GP PCR could be obtained (12). Nig05-A08 clusters with strain LP, although the genetic distance between the two sequences (25% at nucleotide level) is at the upper limit of intralinear diversity (maximum intralinear differences: 21% [II], 28% [III], and 23% [IV]). Therefore, it is not clear if Nig05-A08 represents a new lineage or a sublineage within lineage I. Nevertheless, the sequence of Nig05-A08 provides evidence that the diversity among Lassa virus strains in Nigeria is greater than previously known.

The trees based on the partial sequences further confirm that strains circulating in the southern and central parts of the country cluster with lineages II and III, respectively. The topologies of the trees even suggest a geographical and temporal pattern within the lineages. Taking the GP, NP, and L gene-based phylogenies together (not all strains are represented in every tree), it appears that the age of the branching nodes of lineage II strains correlates with geography. The strains seem to have emerged in the order Owerri (most ancient)-Aba-Abakaliki-Onitsha-Irrua/Ekpoma (most recent), suggesting that the virus spread from southeast to northwest, as shown in Fig. 2. Lineage III strains seem to have been split into two sublineages in the past. One sublineage, represented by the CSF strain, appears to circulate around Jos, while the second

TABLE 2. Phylogenetic analysis of the Lassa virus clade with estimation of substitution rates and divergence times by BEAST^a

Gene	No. of taxa	No. of sites	Demographic model	Mean rate (95% HPD) in substitutions $\times 10^{-4}$ site ⁻¹ yr ⁻¹	Tree likelihood ^d	Age (yr) of MRCA before present median (95% HPD) ^b			
						Lassa (node A)	Lineage II (node B)	Lineage III (node C)	Lineage IV (node D)
GP	82	237	EXP	17.6 (10–25)	–4,458	328 (209–511)	92 (57–142)	180 (119–277)	132 (84–210)
GP	82	237	BSL	17.6 (10–25)	–4,459	334 (197–505)	94 (59–143)	185 (113–275)	132 (81–205)
NP	99	631	EXP	6.7 (4.9–8.7)	–8,979	734 (526–1,015)	226 (167–305)	438 (309–612)	330 (243–454)
NP	99	631	BSL	6.6 (4.7–8.5)	–8,977	747 (528–1,030)	229 (171–321)	453 (314–631)	337 (241–467)
L	34	342	EXP	7.0 (1.9–12)	–3,246	550 (229–1,301) ^c	182 (82–414)	275 (108–656)	280 (130–662)
L	34	342	BSL	5.6 (0.5–10)	–3,246	762 (246–2,915) ^c	222 (70–845)	346 (102–1,277)	350 (125–1,372)

^a Abbreviations: MRCA, most recent common ancestor; 95% HPD, highest posterior density (the smallest interval that contains 95% of the posterior probability density); EXP, exponential growth; BSL, Bayesian skyline. The phylogenetic trees are shown in Fig. 1 right (Nigerian lineages only) and Fig. S2 in the supplemental material (complete trees).

^b The nodes are marked with the respective letter in Fig. 1.

^c Does not include lineage I.

^d Tree likelihood is the sum of the log likelihoods for each site in the alignment and indicates the probability that the sequence data fit with the given tree topology and evolutionary model. The tree likelihood values are nearly identical for the two demographic models (EXP and BSL) for the GP, NP, or L gene, indicating that both models describe the data equally well.

sublineage appears to be distributed in a wider area including Jos, Vom, Zonkwa, Zaria, and Abuja.

The estimated substitution rates for the Lassa virus clade range from 5.6×10^{-4} to 17.6×10^{-4} substitutions \cdot site $^{-1} \cdot$ year $^{-1}$ (Table 2), which is in the same order of magnitude as rate estimations for the prototypic arenavirus lymphocytic choriomeningitis virus (1) and other rodent-borne RNA viruses (14). The rate estimates imply that the ancestors of contemporary Lassa viruses spread through West Africa during the past 300 to 800 years (Table 2). This period appears to be quite short, and thus this result may be interpreted with caution. Prospective studies on virus evolution in the natural reservoir may provide clues as to whether this estimate is realistic.

In conclusion, this study reports genomic sequences of six recent Nigerian strains, representing lineages II and III, and one strain from Guinea. The novel sequences will aid in the design of molecular detection assays for Lassa virus. Phylogenetic analysis of a larger set of partial sequences implies the presence of an additional (sub)lineage of Lassa virus in Nigeria, a directional evolutionary spread of the virus within the country, and a divergence of Lassa virus into the West African lineages during the past 300 to 800 years.

Nucleotide sequence accession numbers. The sequences reported in this paper have been sent to the GenBank database and assigned accession no. GU481056 to GU481079, GU830839, GU979503, to GU979513, HM143866 to HM143868, and HM143870 to HM143881.

The study was supported by a grant from the Alexander von Humboldt Foundation to S.A.O., grant I/82 191 from the Volkswagen Foundation, grant GU 883/1-1 from the German Research Foundation (DFG), FP7 grant 228292 (European Virus Archive) from the European Community, and FP7 Marie-Curie grant PIEF-GA-2009-235164 to E.F.-C. The Department of Virology of the Bernhard Nocht Institute is a WHO Collaborating Centre for Arbovirus and Haemorrhagic Fever Reference and Research (DEU-000115).

ADDENDUM IN PROOF

Partial Sequencing of a further ~200 Lassa virus strains from clinical specimens collected during 2008 through 2010 in Irrua identified one more strain belonging to the lineage/sublineage established by strain Nig05-A08. Both strains are 96% identical in the GP gene fragment amplified by the diagnostic RT-PCR.

REFERENCES

1. Albarino, C. G., et al. 2010. High diversity and ancient common ancestry of lymphocytic choriomeningitis virus. *Emerg. Infect. Dis.* **16**:1093–1100.
2. Atkin, S., et al. 2009. The first case of Lassa fever imported from Mali to the United Kingdom, February 2009. *Euro Surveill.* **14**:2–4.
3. Bowen, M. D., et al. 2000. Genetic diversity among Lassa virus strains. *J. Virol.* **74**:6992–7004.
4. Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**:e88.
5. Ehichioya, D. U., M. Hass, S. Öschlger, et al. 2010. Lassa fever, Nigeria, 2005–2008. *Emerg. Infect. Dis.* **16**:1040–1041.
6. Fichet-Calvet, E., and D. J. Rogers. 2009. Risk maps of Lassa fever in West Africa. *PLoS Negl. Trop. Dis.* **3**:e388.
7. Fisher-Hoch, S. P., et al. 1995. Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. *BMJ* **311**:857–859.
8. Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
9. Günther, S., et al. 2000. Imported lassa fever in Germany: molecular characterization of a new lassa virus strain. *Emerg. Infect. Dis.* **6**:466–476.
10. Lecompte, E., et al. 2006. *Mastomys natalensis* and Lassa fever, West Africa. *Emerg. Infect. Dis.* **12**:1971–1974.
11. Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562–563.
12. Ölschlger, S., et al. 2010. Improved detection of Lassa virus by reverse transcription-PCR targeting the 5' region of S RNA. *J. Clin. Microbiol.* **48**:2009–2013.
13. Omlabu, S. A., et al. 2005. Lassa fever, Nigeria, 2003 and 2004. *Emerg. Infect. Dis.* **11**:1642–1644.
14. Ramsden, C., F. L. Melo, L. M. Figueiredo, E. C. Holmes, and P. M. Zanotto. 2008. High rates of molecular evolution in hantaviruses. *Mol. Biol. Evol.* **25**:1488–1492.
15. Vieth, S., et al. 2007. RT-PCR assay for detection of Lassa virus and related Old World arenaviruses targeting the L gene. *Trans. R. Soc. Trop. Med. Hyg.* **101**:1253–1264.